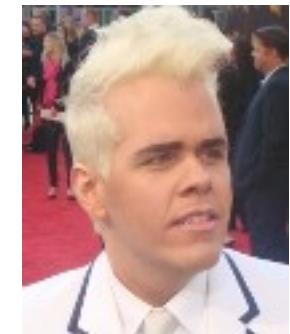
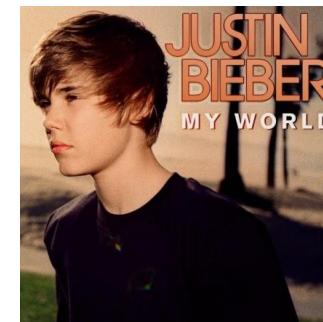


Prevention and Reaction Defending Privacy in the Web 2.0

Michael Hart
Rob Johnson
mhart@cs.stonybrook.edu
Stony Brook University

For all the Web's successes...



...what is the cost to privacy?

Blog Stalkers – Personal Safety for Bloggers

Harsh Words Die Hard on the Web Law Students Feel Lasting Effects of Anonymous Attacks



Hi!

i guess you forgot about adding

me on here?
Firstly, don't flatter yourself. Secondly, you've worked here 5 months and didn't work out that i'm gay? I know i don't prance around the office like a queen, but it's not exactly a secret. Thirdly, that stuff is called your 'job', you know, what i pay you to do. But the fact that you seem able to the simplest of tasks might contribute to how you feel about it. And lastly, you also seem to have forgotten that you have 2 weeks left on your 6 month trial period. Don't bother coming in tomorrow. I'll pop your P45 in the post, and you can come in whenever you like to pick up any stuff you've left here. And yes, i'm serious.

Yesterday at 22:53

Write a comment...

Single Mom Fired From Job For Anonymous Sex Blog

MySpace Predator Caught by Code Details Released About 'Craigslist' Suspect

Updated: March 9, 2009, 5:41 PM ET

f Share 1161

Phoebe Prince: Facebook Bullying Behind Teen's Suicide

Facebooking while out sick gets employee fired

Perspective: I was fired for blogging

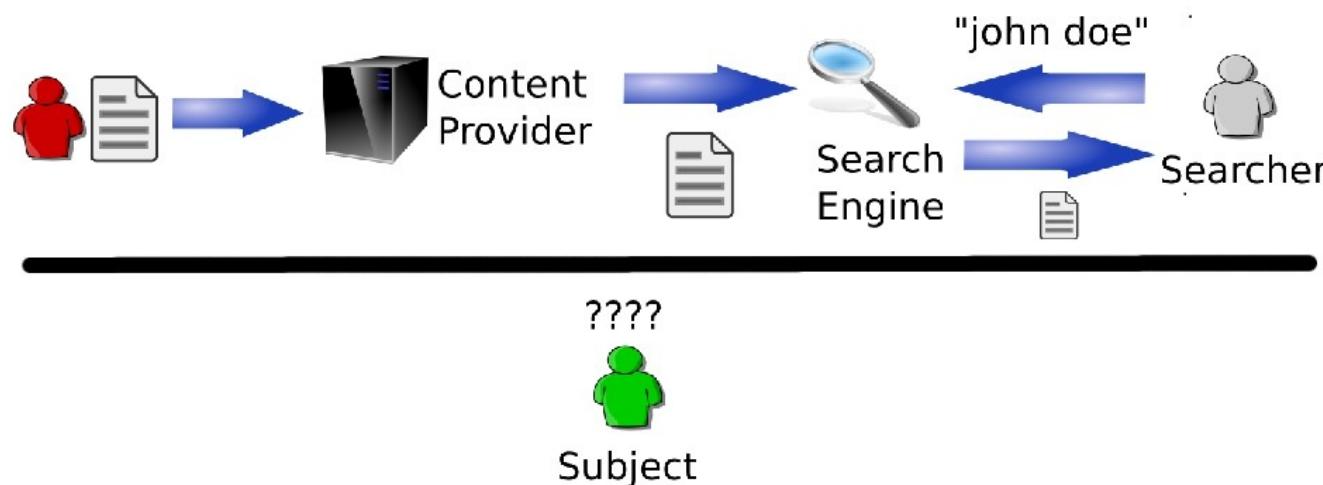
Twitter gets you fired in 140 characters or less

Main sources of privacy invasions

■ Disclosed data



■ Incidental data



What are service providers doing?

■ Disclosed data

- Provide users simplistic access controls

■ Incidental data

Service	Can user make it private?
Facebook	Only if user is tagged
Blogger, LiveJournal, WordPress and other blogging sites	No
MySpace, Hi5, qq, other social networking sites	No
Flickr, Picassa, other photo sharing sites	No
YouTube, MetaCafe and other video sharing sites	No
Other content sharing sites	No

Where these sites come up short

- Privacy controls are too coarse
 - Group permissions by friends or content type
- Lack feedback for actions
 - Users do not know impact of their actions
- No safety net
 - Public by default
- Force users to choose between anonymity and accessibility
 - Who really has 500 best friends?
- Portability

So what do users need?

- Flexibility to encompass all privacy preferences
- Easy to use
 - Users have little patience and time for access control
 - Requires little extra effort
 - Succinct policies for large content collections
- Easy to understand
 - Users know who has access to what
- Safety
 - Infer privacy policy on newly created content

Tag-based privacy policies

- Privacy preferences expressed as rules on tags
 - Only my “college buddies” can see posts marked “Stony Brook University”
- When we have new content
 - Apply rules based on tags to create policy
 - Allow for exception

Why tag-based policies?

- Users already tag the data they post
 - Even on password protected content!
- Tags are extremely flexible
- Enable users to express in familiar terms
 - In terms of their content and attributes
 - Their relationships
 - Both specific (e.g. Emily) and abstract (e.g. co-worker)
- Tag-based policies are portable across services
- Tags are inferable from content
 - Thus, privacy policies are inferable

Do tag-based policies work?

- **Flexible**
 - Subjects wrote policies over disparate sensitive topics
- **Easy to use**
 - Subjects applied tag-based policies significantly faster than an per-item policies
 - Even with over 100 tags to choose
- **Easy to understand**
 - Subjects tag-based policies as accurate as per-item policies
 - Subjects wrote near optimal policies w.r.t. size
 - Result in succinct policies
 - Most privacy policies in less than 5 rules on existing blogs
- **Provides protection**
 - Built a tagger for policy inference that achieved precision and recall over 60% in general case

Incidental data privacy disclosure

- Increasing threat to privacy
 - Sophistication of search engines
 - Integration of real life and the web
- Challenges
 - Incentives
 - Freedom of speech

Responsibility for containment?

????

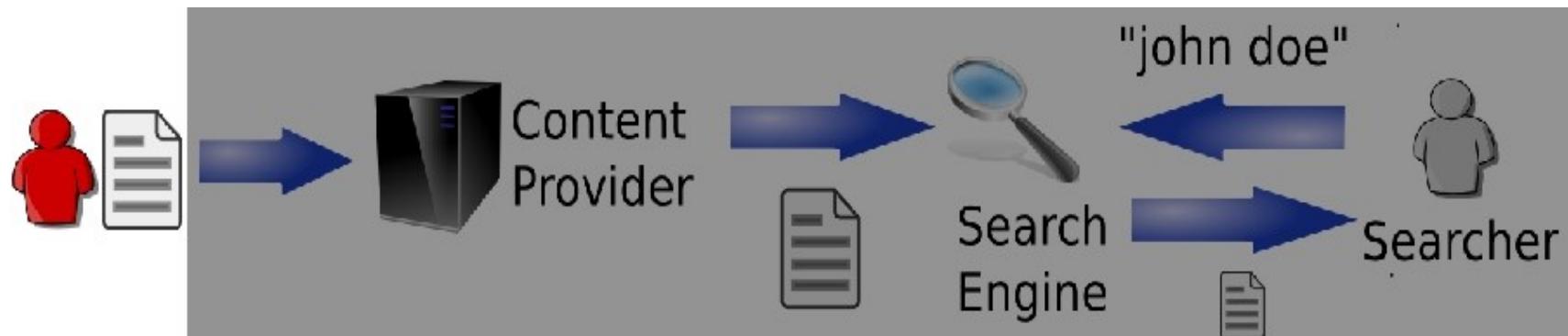


Subject

- The subject of the privacy invasion must contain it
- Options for recourse
 - Litigation
 - Other questionable means
 - Try to influence search engine rankings
 - DoS attack

Who will aide him?

- The content author?

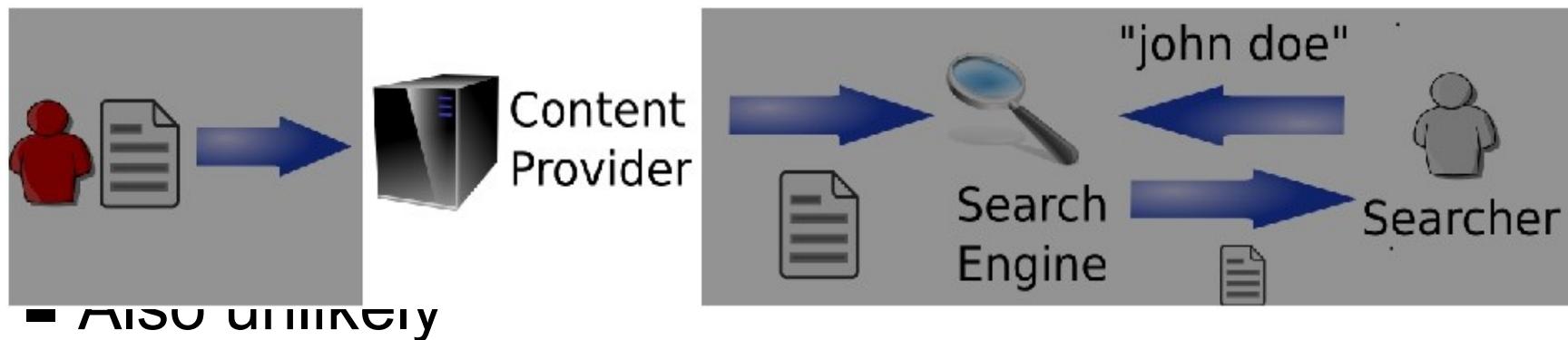


- Unlikely

- Only a few cases of online libel have been prosecuted

Who will aide him?

- The content provider?

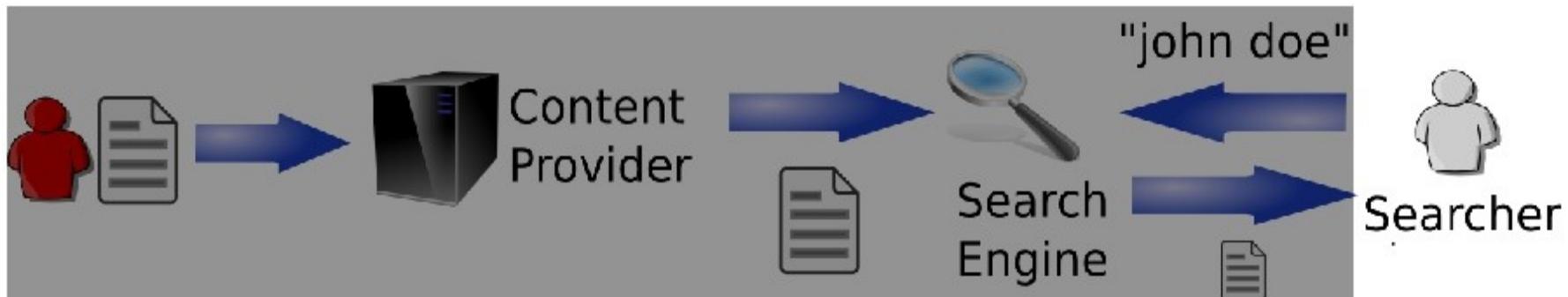


- ~~also uniquely~~

- Goal to serve content, not filter it
- Laws protect them

Who will aide him?

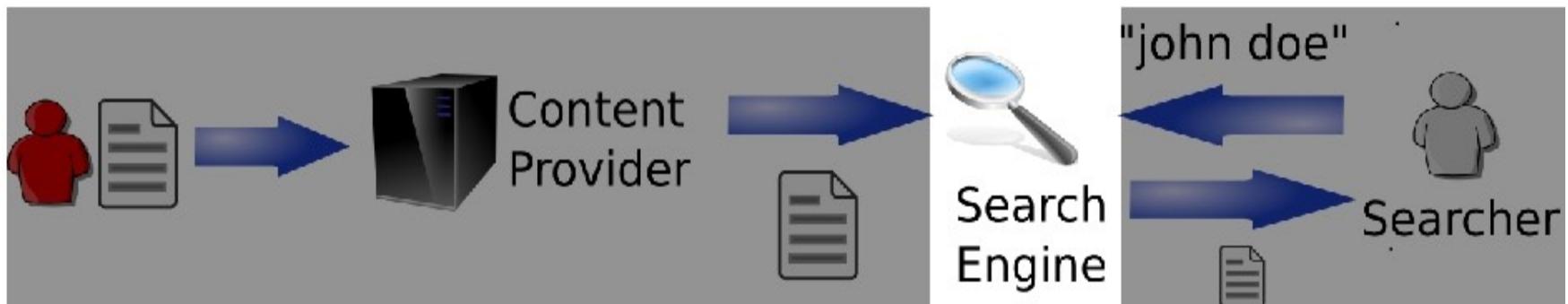
- The Searcher



- A malicious searcher will not
- A friendly searcher cannot

Who will aide him?

■ Search engine



- Improving privacy can improve search results
 - Search for applicant yields work related links

Modifications for people search

- Order results based on
 - Authority
 - Objectivity
- Devalue dubious or opinionated looking sites
 - Identify unmoderated forums
 - Sorry Auto-Admit, 4Chan and Juicy Campus
- Display ratings beside result:
 - Neutrality
 - Factuality

More ambitious features

- Require more specific search queries
 - Searcher demonstrates some knowledge of existence of relationship
- Allow users to express privacy preferences
 - Search engine can factor user preference into search results
 - Users declare personal/private topics
 - What's fair game
 - Search engines (may) apply to search results
 - More “questionable” the results, more influence

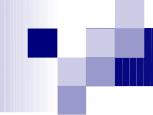
The larger picture

- How do we help the user?
- Usability!
- Inspire better access control
- Knowledge is the key to the kingdom



Parting thoughts

- Privacy for disclosed data
 - Deploy tag-based privacy policies
 - Use ML and NLP to automate privacy
- Privacy for incidental data
 - Don't censor
 - Steer users away from privacy invasive material
 - May improve search results
 - Preserve free speech rights



Thanks!

Questions?

mhart@cs.stonybrook.edu